WEIBULL DISTRIBUTION PARAMETER ESTIMATION USING SPREADSHEETS

ABSTRACT

A spreadsheet is developed for modeling wind distribution through the Weibull distribution. Two parameters of the Weibull distribution are sufficient for the modeling. The excel macro has been used to determine these parameters. The parameters are determined by six different methods: graphical method (GM), maximum likelihood method (MLM), method of moments (MM), energy pattern factor method (EPFM), empirical method (EM), and modified maximum likelihood method (MMLM). Each of these six methods has a corresponding button on the spreadsheet; this button calculates parameters '*k*' and '*c*' of the Weibull distribution on activation. These buttons also find the average of wind speed and statistical errors; root mean square error (RMSE), chi-square statistics (χ^2), and the coefficient of determination for validation of results. To check the effectiveness of the spreadsheet, the wind speed data of Volkel, Gilze-Rijen, Eindhoven, and Herwijnen, cities in the Netherlands, during the years 2015 and 2016, were used. The data were obtained from the Royal Netherlands Meteorological Institute. The spreadsheet also generates the probability density function using parameter values obtained by the six methods; these are shown on a single plot of the probabilities against the wind speed.

1 INTRODUCTION

Spreadsheets allow students to gain experience working with real data. Spreadsheets make theoretical or complex models accessible by providing real-world examples and allowing hypothetical analyses. Tables on a printed page are stale, while spreadsheet representations allow students to interact with the underlying concepts. Instructors can use Excel to convert any spreadsheet into a web page and share it with students, which is an effective way of increasing teaching efficiency.

In this paper, we developed a worksheet to calculate the Weibull parameters and other parameters, such as σ, $\bar{v}$, etc., for the six numerical methods discussed below so that we can compare these methods to find the most efficient method for adjusting the Weibull distribution of wind speeds.

1.1 Location and Climate

Volkel is in the North Brabant province in the southern Netherlands. Volkel Air Base is a military airbase used by the Royal Netherlands Air Force (RNLAF) and is located near the village of Volkel in North Brabant, Netherlands. Gilze-Rijen is a municipality, in the south of the Netherlands, which contains four villages Rijen, Gilze, Hulten and Molenschot [1]. Eindhoven is also a municipality and a city in the southern Netherlands [2]. Eindhoven has an oceanic climate with slightly warmer summers and colder winters than the coastal parts of the Netherlands. Herwijnen, a village in the Dutch province of Gelderland, is a part of the Lingewaal municipality, and located nearly eleven kilometers east of Gorinchem.

1.2 Weibull Distribution

To evaluate the wind energy in a city, it is essential to make extensive climatic observations in that area. The wind speed is a significant random variable which affects the most accurate results on the energy potential of the site. The wind speed data can be represented by a probability density function (PDF). Currently the most extensively used tool to determine the wind potential is the Weibull distribution.

2 THE WEIBULL PARAMETERS DETERMINATION METHODS

Because wind speed is a random variable, statistical analysis is required to determine a region's wind potential [3,4,5,6]. This requires the wind speed data as a function of time. The Weibull distribution can be defined as a probability density function *f(v)* and a cumulative distribution function *F(v)*, for the given wind speed data, determined by the following equations: [7,8,9,10]

$$f(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} e^{-(v/c)^k} \tag{1}$$

$$F(v) = 1 - e^{-(v/c)^k} \tag{2}$$

where *k* and *c* are the Weibull parameters, and *v* is the wind speed.

We have used the following six methods to estimate the Weibull parameters:

2.1 Graphical method

The first method is the graphical method (GM) which is obtained through the cumulative distribution function *F(v)*. In this method the wind speed data are interpolated by a straight line, utilizing the idea of least square approximation. The double logarithmic transformation equation for this method can be represented as [11]:

$$\ln[-\ln\{1 - F(v)\}] = k \ln\frac{v}{c} \tag{3}$$

2.2 Maximum likelihood method

The second is the maximum likelihood method (MLHM). In this method numerical iterations are required to determine the Weibull parameters [12]. The Weibull parameters '*k*' and '*c*' are obtained using the following equations [13]:

$$k = \left[\frac{\sum_{i=1}^{n} v_i^k \ln v_i}{\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln v_i}{n}\right]^{-1} \tag{4}$$

$$c = \left(\frac{1}{n}\sum_{i=1}^{n} v_i^k\right)^{\frac{1}{k}} \tag{5}$$

where '*n*' is the number of observations, and 'v_i' is the measured wind speed in the interval '*i*'.

2.3 Energy pattern factor method

The third is the energy pattern factor method (EPFM). This method is related to the mean data of wind speed and is defined by the equations given below [14]:

$$E_{pf} = \frac{\overline{v^3}}{\bar{v}^3} \tag{6}$$

$$k = 1 + \frac{3.69}{(E_{pf})^2} \quad (7)$$

$$\bar{v} = c\Gamma(1 + 1/k) \quad (8)$$

where 'E_{pf}' is the energy pattern factor, and the gamma function defined by the following equation:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \quad (9)$$

2.4 Moment method

The fourth method is called the moment method (MM). This method can be used as a substitute for the maximum likelihood method (MLHM) [15]. In this case the Weibull parameters are determined by the equations:

$$\bar{v} = c\Gamma(1 + 1/k) \quad (10)$$

$$\sigma = c[\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)]^{1/2} \quad (11)$$

where $\bar{v}$ is the average wind speed, and σ is the standard deviation of the collected wind speed data.

2.5 Empirical method

This method can be considered as a special case of the moment method (MM) in which the Weibull parameters are determined by equations [11]:

$$k = \left(\frac{\sigma}{\bar{v}}\right)^{-1.086} \quad (12)$$

$$\bar{v} = c\Gamma(1 + 1/k) \quad (13)$$

2.6 Modified maximum likelihood method

The most efficient method is the modified maximum likelihood method (MMLHM), but this method can only be used if the available data are already given in the shape of the Weibull distribution. It also requires numerical iterations, as in the maximum likelihood method, to solve the following equations [11]:

$$k = \left[\frac{\sum_{i=1}^{n} v_i^k \ln(v_i) f(v_i)}{\sum_{i=1}^{n} v_i^k f(v_i)} - \frac{\sum_{i=1}^{n} \ln(v_i) f(v_i)}{f(v \geq 0)}\right]^{-1} \quad (14)$$

$$c = \left(\frac{1}{f(v \geq 0)} \sum_{i=1}^{n} v_i^k f(v_i)\right)^{\frac{1}{k}} \quad (15)$$

where $f(v_i)$ represents the Weibull frequency, and $f(v \geq 0)$ is the probability of wind speed ≥ 0.

3 STATISTICAL ANALYSIS

The efficiency of all these methods is analyzed by performing three tests, namely RMSE, χ^2, and R^2. These tests are defined as follows:

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{n}(y_i - x_i)^2\right)^{\frac{1}{2}} \quad (16)$$

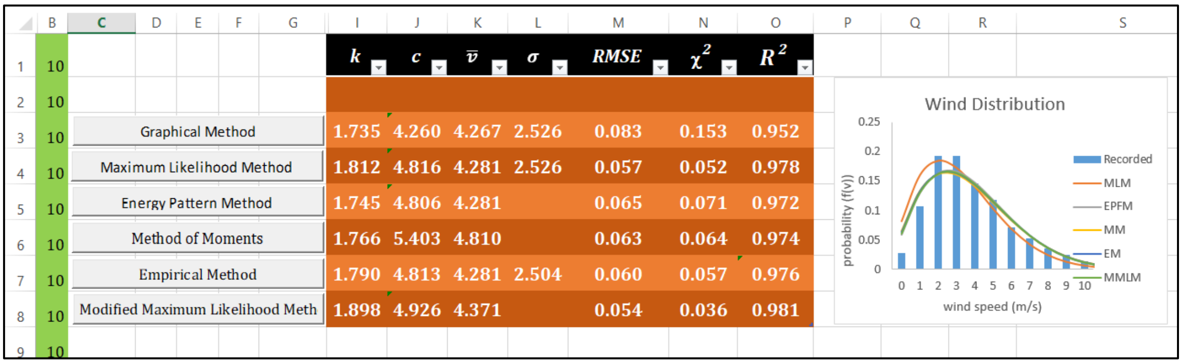
$$\chi^2 = \frac{1}{(N-n)}\sum_{i=1}^{n}(y_i - x_i)^2 \quad (17)$$

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - z_i)^2 - \sum_{i=1}^{N}(y_i - x_i)^2}{\sum_{i=1}^{N}(y_i - z_i)^2} \quad (18)$$

where 'N' is the number of observations, 'y_i' is the frequency of observations, 'x_i' is the frequency of Weibull, 'z_i' is the mean wind speed, and 'n' is the number of constants used.

4 RESULTS

Figure 1 shows a screenshot of the spreadsheet for modeling wind distribution. The spreadsheets show six different buttons for six different methods for calculating Weibull parameters. The spreadsheet also shows parameters 'k' and 'c' and the errors calculated by each method. The plot on the spreadsheet is generated by the 'k' and 'c' values determined by six methods. Figures 2, 4, 6, and 8 show the Weibull distribution of the wind speed data, i.e., these figures give the plot probability function *f(v)* versus the wind speed for Volkel, Gilze-Rigen, Eindhoven, and Herwignen, respectively for the year *2015*. Similarly, figures 3, 5, 7, & 9 give the same for *2016*. Figures 2 to 9 show the wind speed distribution histogram and Weibull pdf for wind speed data obtained by six different methods: graphical, method of moments, maximum likelihood method, modified maximum likelihood method, energy pattern factor, and empirical method. The Weibull parameters, 'k' and 'c', the mean wind speed '$\bar{v}$', and the standard deviation 'σ' are shown in Tables 1 to 8. The estimated 'k' and 'c' are compared graphically for each numerical method and all four cities, for *2015* and *2016*, in Figures 10 to 13. The statistical errors (root mean square error and chi-square) are also compared graphically for the four cities and six methods; see Figures 14 to 19. The coefficient of determination is also compared in Figures 18 & 19.



	k	c	$\bar{v}$	σ	RMSE	χ^2	R^2
Graphical Method	1.735	4.260	4.267	2.526	0.083	0.153	0.952
Maximum Likelihood Method	1.812	4.816	4.281	2.526	0.057	0.052	0.978
Energy Pattern Method	1.745	4.806	4.281		0.065	0.071	0.972
Method of Moments	1.766	5.403	4.810		0.063	0.064	0.974
Empirical Method	1.790	4.813	4.281	2.504	0.060	0.057	0.976
Modified Maximum Likelihood Meth	1.898	4.926	4.371		0.054	0.036	0.981

Figure 1 *Screenshot of the Spreadsheet*

5 DISCUSSION

Among the six numerical methods discussed above, the ranking is as follows. The modified maximum likelihood method gives the best curve fit. The second best is the empirical method; the remaining are in the following order: moment method, energy pattern factor method, and the graphical method. The above statement is supported by the statistical errors RMSE, χ^2 and R^2 calculated for the cities Volkel, Gilze-Rijen, Eindhoven, and Herwijnen.

The values of k determined from the moment method, empirical method, and the modified maximum likelihood method are very close. In contrast, those determined from the graphical method, energy pattern factor method, and the maximum likelihood method significantly differ. The moment method gives the maximum value of k except in the case of Volkel for the year 2016.

It is observed that the estimated 'c' from the energy pattern factor method, moment method, empirical method, and in some cases, the modified maximum likelihood method, have closer values, while the values estimated from the graphical method, and the maximum likelihood method slightly differ. The maximum values of c occurred for the Herwijnen in 2015.

6 CONCLUSIONS

A spreadsheet was developed on excel to model wind distribution. The spreadsheet is easily accessible to everyone, so this simple GUI would help to calculate Weibull parameters using six different methods. It allows students to organize, calculate, graph, and analyze data efficiently.

Among the six numerical methods, the modified maximum likelihood method is efficient for determining the parameters k and c to fit Weibull distribution curves for wind

speed data. In contrast, the graphical method is the least effective method to fit Wweibull distribution curves.

The maximum value of k is found to be 2.4 from the moment method, and the maximum value of c is 5.2146 m/s from the modified maximum likelihood method. The mean wind speeds estimated for all four stations range from 4.371 m/s to 4.602 m/s in 2015 and from 3.992 m/s to 4.183 m/s in 2016. The standard deviation lies between 2.266 m/s and 2.551 m/s in 2015, and between 2.021 m/s and 2.213 m/s in 2016.

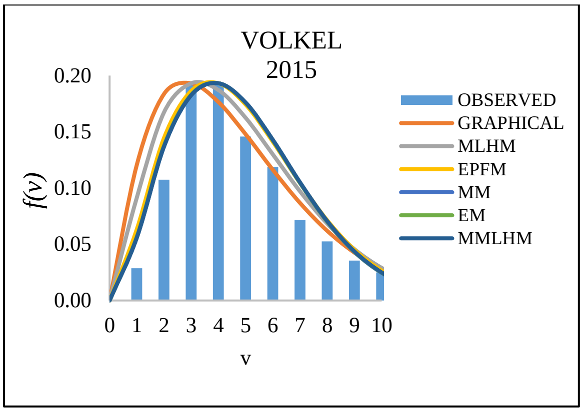


Figure 2 *Weibull distribution-Volkel, year 2015*

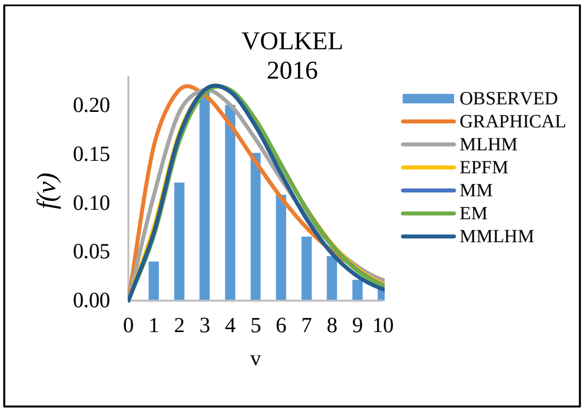


Figure 3 *Weibull distribution-Volkel, year 2016*

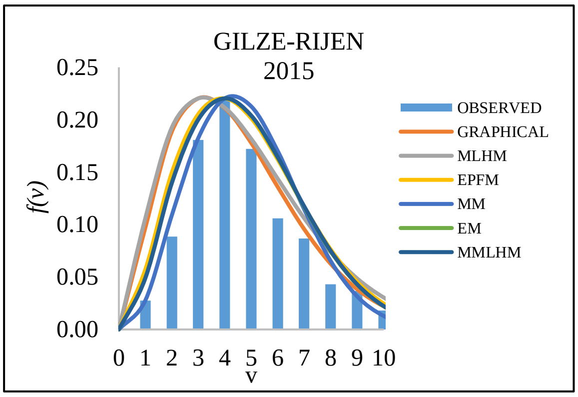


Figure 4 *Weibull distribution-Gilze-Rijen, year 2015*

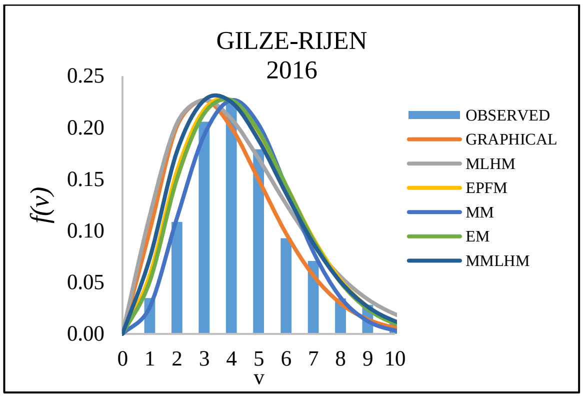


Figure 5 *Weibull distribution-Gilze-Rijen, year 2016*

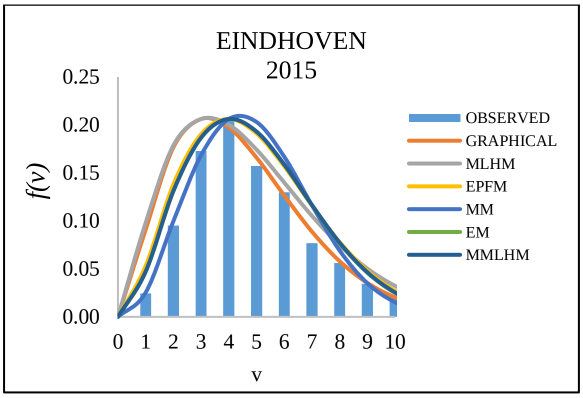


Figure 6 *Weibull distribution-Eindhoven, year 2015*

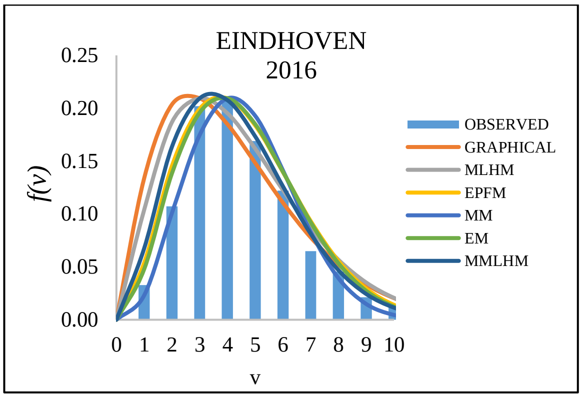


Figure 7 *Weibull distribution-Eindhoven, year 2016*

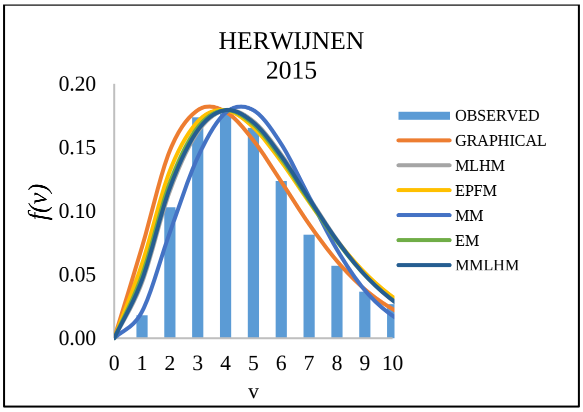


Figure 8 *Weibull distribution-Herwijnen, year 2015*

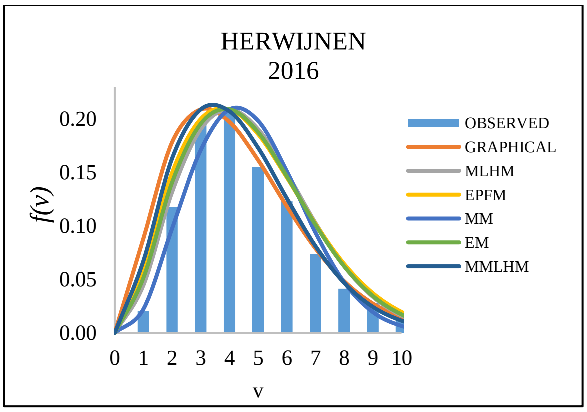


Figure 9 *Weibull distribution-Herwijnen, year 2016*

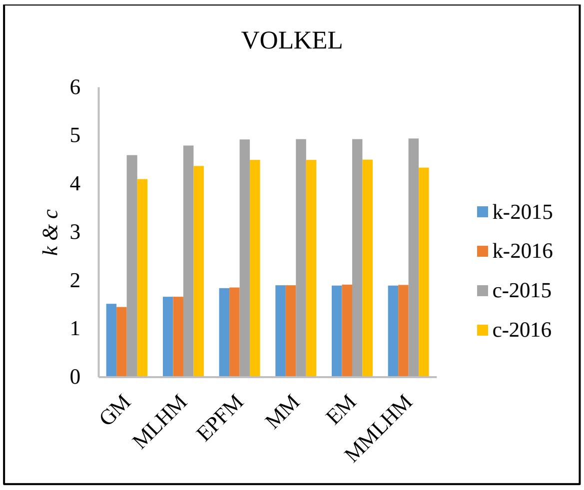


Figure 10 *k and c - Volkel - 2015 & 2016*

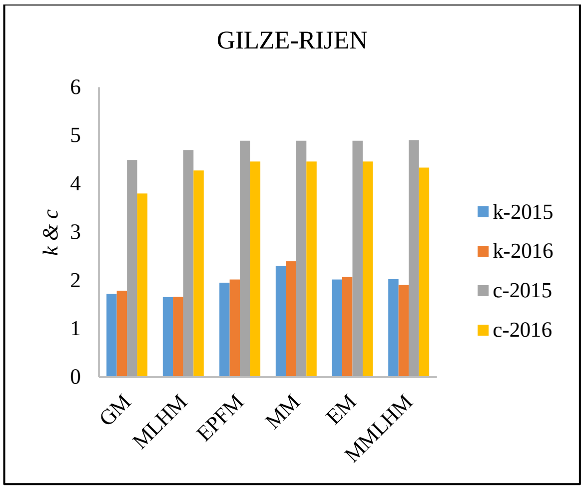


Figure 11 *k and c – Gilze-Rijen - 2015 & 2016*

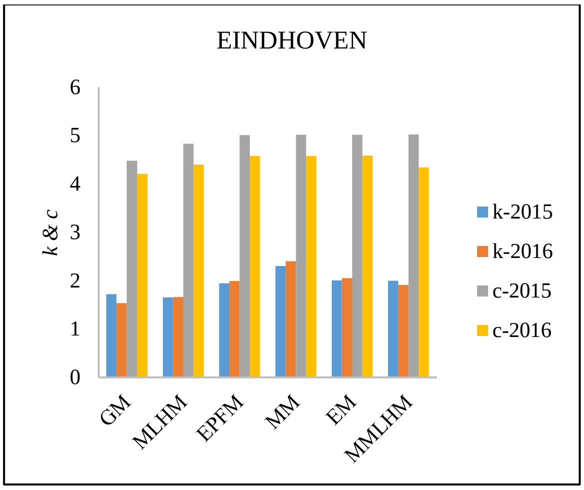


Figure 12 *k and c - Eindhoven - 2015 & 2016*

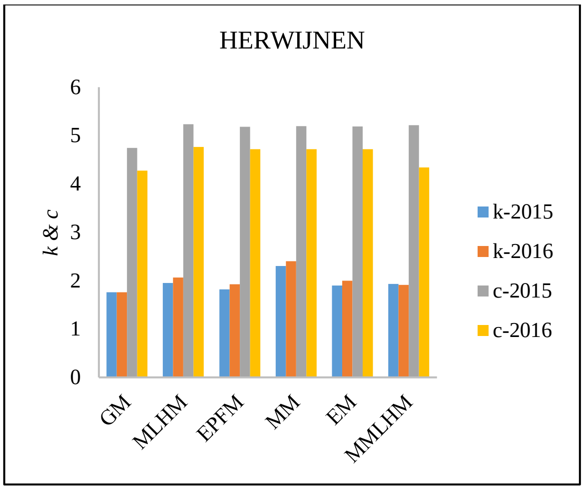


Figure 13 *k and c - Herwijnen - 2015 & 2016*

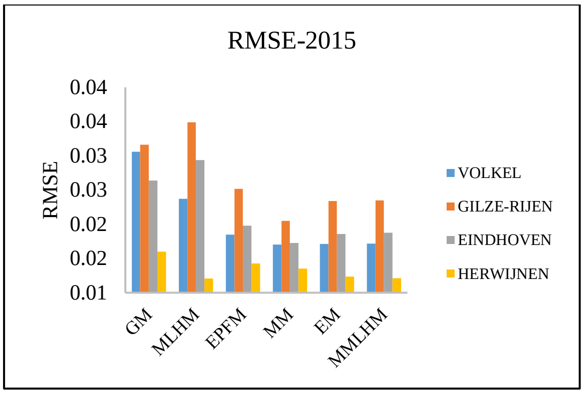


Figure 14 *RMSE - 2015*

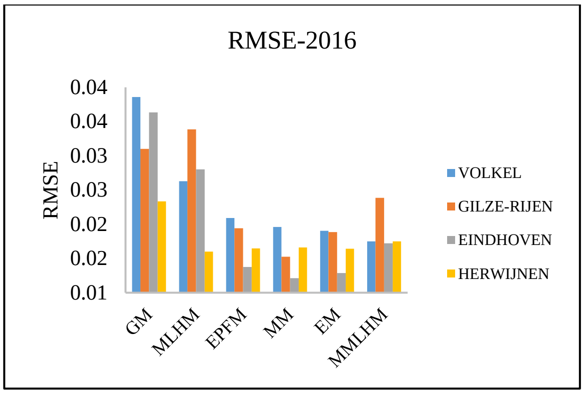


Figure 15 *RMSE - 2016*

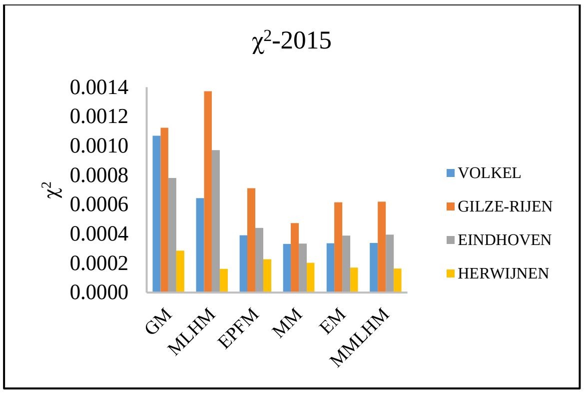


Figure 16 *χ^2- 2015*

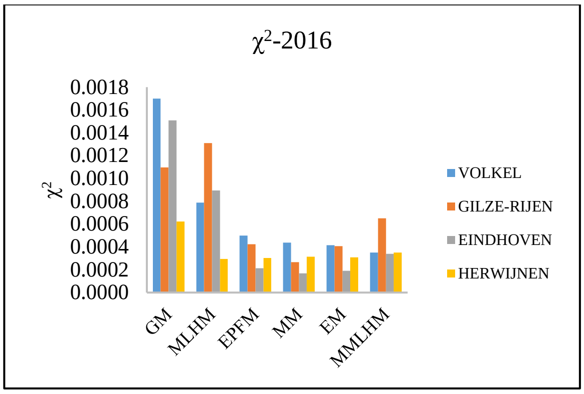


Figure 17 *χ^2- 2016*

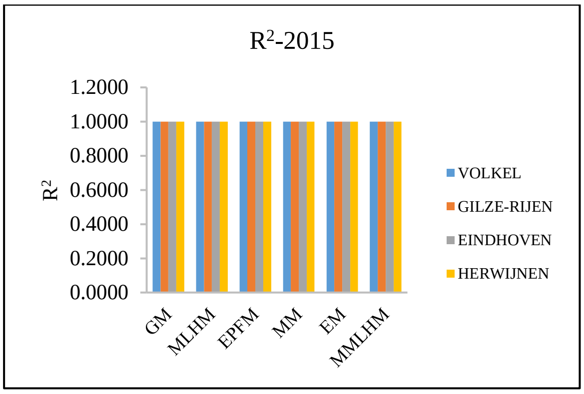


Figure 18 *R^2- 2015*

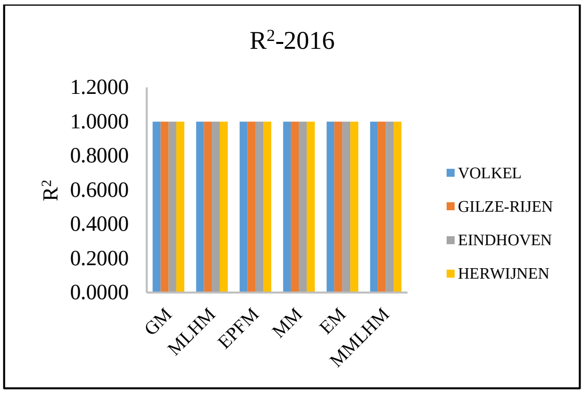


Figure 19 *R^2- 2016*

Table 1 *STATISTICAL ANALYSIS: VOLKEL, year 2015*

METHOD	2015						
	k	*c*	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.5181	4.5951	0.0306	0.0011	1.0000	0.9158	0.2241
Max. Likelihood	1.6613	4.7911	0.0237	0.0006	1.0000	0.9592	0.4201
Energy Pattern Factor	1.8403	4.9201	0.0185	0.0004	1.0000	0.9788	0.5491
Moment	1.9000	4.9258	0.0170	0.0003	1.0000	0.9807	0.5548
Empirical	1.8936	4.9253	0.0171	0.0003	1.0000	0.9806	0.5543
Mod. Max. Likelihood	1.8978	4.9367	0.0172	0.0003	1.0000	0.9804	0.5657

Table 2 *STATISTICAL ANALYSIS: VOLKEL, year 2016*

METHOD	2016						
	k	*c*	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.4520	4.0970	0.0386	0.0017	1.0000	0.8539	0.1055
Max. Likelihood	1.6613	4.3681	0.0263	0.0008	1.0000	0.9429	0.3765
Energy Pattern Factor	1.8581	4.4947	0.0209	0.0005	1.0000	0.9780	0.5031
Moment	1.9000	4.4982	0.0196	0.0004	1.0000	0.9814	0.5067
Empirical	1.9151	4.4993	0.0191	0.0004	1.0000	0.9825	0.5078
Mod. Max. Likelihood	1.9102	4.3378	0.0175	0.0004	1.0000	0.9867	0.3463

Table 3 *STATISTICAL ANALYSIS: GILZE-RIJEN, year 2015*

METHOD	2015						
	k	*c*	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.7209	4.4933	0.0316	0.0011	1.0000	0.9394	0.1578
Max. Likelihood	1.6563	4.7029	0.0349	0.0014	1.0000	0.9342	0.3674
Energy Pattern Factor	1.9540	4.8897	0.0252	0.0007	1.0000	0.9747	0.5542
Moment	2.3000	4.8938	0.0205	0.0005	1.0000	0.9777	0.5583
Empirical	2.0229	4.8930	0.0234	0.0006	1.0000	0.9776	0.5575
Mod. Max. Likelihood	2.0274	4.9055	0.0235	0.0006	1.0000	0.9775	0.5700

Table 4 *STATISTICAL ANALYSIS: GILZE-RIJEN, year 2016*

METHOD	2016						
	k	*c*	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.7861	3.7999	0.0310	0.0011	1.0000	0.8861	0.1551
Max. Likelihood	1.6613	4.2801	0.0339	0.0013	1.0000	0.9054	0.3251
Energy Pattern Factor	2.0186	4.4635	0.0194	0.0004	1.0000	0.9748	0.5084
Moment	2.4000	4.4615	0.0153	0.0003	1.0000	0.9873	0.5065
Empirical	2.0736	4.4650	0.0188	0.0004	1.0000	0.9788	0.5100
Mod. Max. Likelihood	1.9102	4.3378	0.0239	0.0007	1.0000	0.9736	0.3828

__Table 5__ STATISTICAL ANALYSIS: EINDHOVEN, year 2015

METHOD	2015						
	k	c	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.7221	4.4787	0.0264	0.0008	1.0000	0.9468	0.0352
Max. Likelihood	1.6563	4.8275	0.0294	0.0010	1.0000	0.9487	0.3840
Energy Pattern Factor	1.9474	5.0110	0.0198	0.0004	1.0000	0.9841	0.5676
Moment	2.3000	5.0157	0.0172	0.0003	1.0000	0.9820	0.5722
Empirical	2.0049	5.0142	0.0186	0.0004	1.0000	0.9858	0.5707
Mod. Max. Likelihood	2.0012	5.0224	0.0187	0.0004	1.0000	0.9856	0.5789

__Table 6__ STATISTICAL ANALYSIS: EINDHOVEN, year 2016

METHOD	2016						
	k	c	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.5342	4.2063	0.0363	0.0015	1.0000	0.8634	0.1453
Max. Likelihood	1.6613	4.4023	0.0280	0.0009	1.0000	0.9217	0.3413
Energy Pattern Factor	1.9930	4.5821	0.0138	0.0002	1.0000	0.9829	0.5210
Moment	2.4000	4.5811	0.0121	0.0002	1.0000	0.9902	0.5200
Empirical	2.0486	4.5841	0.0129	0.0002	1.0000	0.9863	0.5230
Mod. Max. Likelihood	1.9102	4.3378	0.0172	0.0003	1.0000	0.9802	0.2768

__Table 7__ STATISTICAL ANALYSIS: HERWIJNEN, year 2015

METHOD	2015						
	k	c	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.7580	4.7419	0.0160	0.0003	1.0000	0.9718	0.1387
Max. Likelihood	1.9522	5.2315	0.0121	0.0002	1.0000	0.9866	0.6283
Energy Pattern Factor	1.8191	5.1788	0.0143	0.0002	1.0000	0.9826	0.5756
Moment	2.3000	5.1960	0.0135	0.0002	1.0000	0.9777	0.5928
Empirical	1.8983	5.1874	0.0124	0.0002	1.0000	0.9864	0.5842
Mod. Max. Likelihood	1.9314	5.2146	0.0121	0.0002	1.0000	0.9867	0.6114

__Table 8__ STATISTICAL ANALYSIS: HERWIJNEN, year 2016

METHOD	2016						
	k	c	*RMSE*	χ^2	R^2	*CORREL.*	$\lvert c-\bar{v}\rvert$
Graphical	1.7621	4.2778	0.0233	0.0006	1.0000	0.9396	0.0946
Max. Likelihood	2.0616	4.7656	0.0160	0.0003	1.0000	0.9925	0.5824
Energy Pattern Factor	1.9278	4.7162	0.0165	0.0003	1.0000	0.9854	0.5330
Moment	2.4000	4.7188	0.0166	0.0003	1.0000	0.9885	0.5357
Empirical	1.9969	4.7201	0.0164	0.0003	1.0000	0.9898	0.5369
Mod. Max. Likelihood	1.9102	4.3378	0.0175	0.0003	1.0000	0.9799	0.1547

__Table 9__ Mean wind speed and standard deviation

STATION	2015		2016	
	Mean wind speed $\bar{v}$ (m/s)	Standard deviation σ (m/s)	Mean wind speed $\bar{v}$ (m/s)	Standard deviation σ (m/s)
VOLKEL	4.3710	2.4280	3.992	2.194
GILZE-RIJEN	4.3355	2.2661	3.955	2.021
EINDHOVEN	4.4435	2.3419	4.061	2.098
HERWIJNEN	4.6032	2.5511	4.183	2.213

REFERENCES

[1] Retrieved from: http://www.weatherbase.com/weather/weathersummary.php3?s=5360&cityname=Gilze%2C+Netherlands&units=

[2] Schippers, J. L. (2009). Stadsrechten en een stadswapen, 1232: het vroege Eindhoven. *De Canon van Eindhoven*, 29.

[3] Celik, A. N. (2004). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey. *Renewable Energy*, *29*(4), 593-604.

[4] Akpinar, E. K., & Akpinar, S. (2004). Determination of the wind energy potential for Maden-Elazig, Turkey. *Energy Conversion and Management*, *45*(18-19), 2901-2914.

[5] Deaves, D. M., & Lines, I. G. (1997). On the fitting of low mean windspeed data to the Weibull distribution. *Journal of Wind Engineering and Industrial Aerodynamics*, *66*(3), 169-178.

[6] Celik, A. N., Makkawi, A., & Muneer, T. (2010). Critical evaluation of wind speed frequency distribution functions. *Journal of Renewable and Sustainable Energy*, *2*(1), 013102.

[7] Ohunakin, O. S., Adaramola, M. S., & Oyewola, O. M. (2011). Wind energy evaluation for electricity generation using WECS in seven selected locations in Nigeria. *Applied Energy*, *88*(9), 3197-3206.

[8] Ucar, A., & Balo, F. (2009). Evaluation of wind energy potential and electricity generation at six locations in Turkey. *Applied Energy, 86*(10), 1864-1872.

[9] Chang, T. P. (2011). Estimation of wind energy potential using different probability density functions. *Applied Energy*, *88*(5), 1848-1856.

[10] Kwon, S. D. (2010). Uncertainty analysis of wind energy potential assessment. *Applied Energy*, *87*(3), 856-865.

[11] Chang, T. P. (2011). Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy*, *88*(1), 272-282.

[12] Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, *10*(4), 507-521.

[13] Stevens, M. J. M., & Smulders, P. T. (1979). The estimation of the parameters of the Weibull wind speed distribution for wind energy utilization purposes. *Wind Engineering*, 132-145.

[14] Akdağ, S. A., & Dinler, A. (2009). A new method to estimate Weibull parameters for wind energy applications. *Energy Conversion and Management*, *50*(7), 1761-1766.

[15] Justus, C. G., & Mikhail, A. (1976). Height variation of wind speed and wind distributions statistics. *Geophysical Research Letters*, *3*(5), 261-264.